# Investigating the dynamics of human skeletal troponin C: A computational study

Undergraduate Research Thesis

Presented in partial fulfillment of the requirements for graduation with honors research distinction in Biochemistry in the undergraduate colleges of The Ohio State University

by Matt Belardo

The Ohio State University April 2017

Project Advisor: Doctor Steffen Lindert, Department of Chemistry and Biochemistry

**Acknowledgements**

I would like to thank Dr. Steffen Lindert for his guidance throughout the research process, and the entire Lindert research group for their helpful discussions on troponin and computation

Table of Contents:

# Chapter 1:

# Troponin and its Role in Muscle Contraction

Muscle contraction is one of the most basic aspects of human life. It controls essentially all types of movement of or within the body. The muscles of the heart contract and send their nutrient-rich cargo to the farthest reaches of the body. The diaphragm rises and falls, allowing air to rush in and out of the lungs. And of course, the skeletal muscles pull on nearly every bone in the body, somehow creating the fluid motions that our bodies make every day.

Being so crucial to advanced life, it has also been a significant topic of study for researchers. The study of skeletal muscle contraction in particular, has been extensive, both its physiology and its biochemistry[1]. The skeletal muscle unit of contraction, the sarcomere, is composed of two filamentous chains of proteins, one thin and one thick, as shown in Figure 1.[1] Upon signaling from the nervous system, the myosin chains pull the actin chains towards their center. The increased overlap of the two fibers leads to an overall shortening of the muscle fiber, resulting in contraction[2].

At the microscopic level, this process is fairly intricate. Myosin “heads”- globular domains protruding from the protein chain- bind to specific sites on the actin molecules. Upon binding, these heads undergo a “flexing” motion: the length of protein connecting the head to the main protein body bends, pulling the actin chain along with it, utilizing the release of a free phosphate from hydrolyzed ATP[3]. The subsequent binding of a new ATP molecule allows the myosin head to release from its binding site on the actin chain and resume its original position to allow for binding of another site[3].

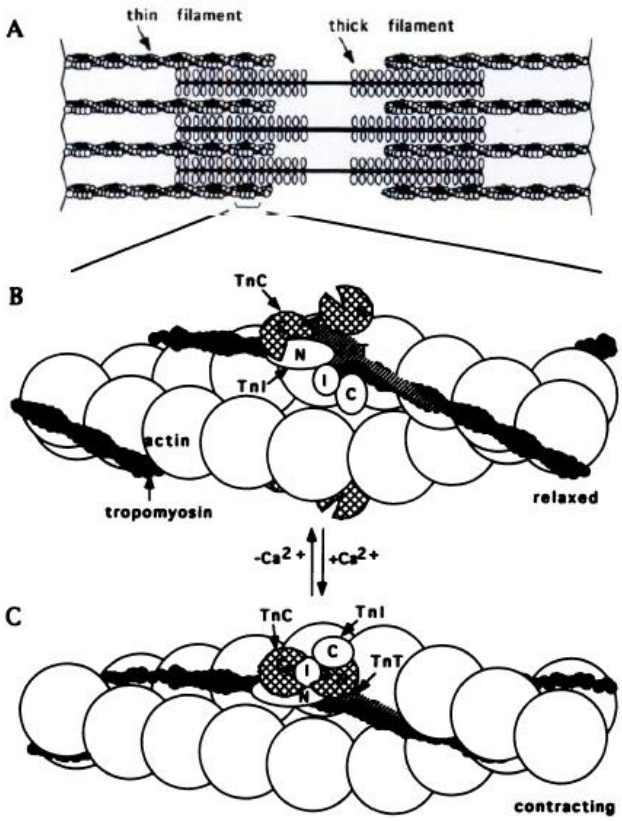



Figure 1: A) Model of the sarcomere, the fundamental unit of contraction in skeletal muscle B) Model of the actin-tropomyosin-troponin complex in the relaxed position C) Model of the complex during contraction[1]

Obviously, this process must be highly regulated. Since skeletal muscles work in pairs[3], the body must be able to activate and deactivate contraction relatively quickly to allow for fluid movements. In skeletal muscles, this regulation is achieved through use of the troponin-tropomyosin complex[4]. The binding sites on actin are blocked by tropomyosin under normal conditions. Without access to these sites, myosin cannot move the thin filament and contract the muscle fiber[5].

Troponin complex: Structure

The troponin complex consists of three distinct subunits. Troponin T (TnT) serves a primarily structural purpose [1], binding to tropomyosin as well as the other subunits of troponin to affix the protein and form the troponin-tropomyosin complex. Troponin I (TnI) is the inhibitory unit of troponin, and it inhibits the ATPase activity of the actomyosin complex in a $Ca^{2+}$-independent manner[6]. The binding of this subunit to actin is significantly increased by the presence of tropomyosin, as is its inhibitory activity[7]. While these subunits are vital to the overall contractile activity of skeletal muscle, they are of only secondary concern in this investigation.

The primary focus of this study is troponin C (TnC), the regulatory calcium-binding subunit of the complex[1]. TnC contains two globular domains connected by a long helix[8]. Each domain is composed of two helix-loop-helix EF hand motifs, each of which contains a calcium-binding site in the loop. The C-terminal domain contains sites that bind calcium with relatively high affinity, but these loops also bind magnesium, which will occupy the sites under physiological conditions due to its greater concentration in muscle cells[4, 7, 9]. The N-terminal domain, on the other hand, has sites that bind with much lower affinity, but increased specificity. This makes the domain much more effective for regulatory purposes than the C-terminal domain, whose function is primarily structural[10, 11].

Troponin C and the calcium switch

Intake of calcium into the muscle cell induces the contraction event within the sarcomere[2]. In a model first proposed by Herzberg et al.[12], calcium binding causes the N-terminal domain of TnC to adopt a conformation that exposes a normally-buried

hydrophobic patch. Various site directed mutant TnC proteins designed to stabilize or destabilize the apo or calcium-bound forms seem to support this model, demonstrating that limiting the molecule's ability to adopt different conformations affected both its calcium affinity and its regulatory activity[13-15]

This activity is responsible for removing the previously mentioned inhibition of the actomyosin activity by TnI[16]. These two subunits bind together at a number of sites[9], and their affinity is significantly increased in the presence of calcium, though it is not entirely removed in its absence[17]. It follows that the removal of inhibition is related to this change in binding rather than a total dissociation of the two subunits.

Given this change in TnC-TnI binding, as well as TnI's binding and inhibition of actomyosin activity, it seems that TnI movement acts as a physical switch that allows activation and deactivation of muscle contraction[18]. This switch in conformation leads to a shift in the tropomyosin conformation, exposing the myosin-binding sites on actin and allowing the contractile activity of myosin to begin[5].

The initial step—calcium binding and the subsequent conformational shift in the N-terminal region of troponin C—is the primary focus of this investigation. The purely structural nature of this particular mechanism makes it an ideal candidate for computational studies. Given the C-terminal region's lack of regulatory function and the computational load of larger molecules, only the N-terminal region of troponin C is utilized in the research methods that will be detailed in Chapters 2 and 3.

Computational Research

This study is not the first time that computation has been used in the study of troponin activity. A number of publications have utilized simulations of troponin in

various conditions to elucidate information about the mechanics of this opening action, both in skeletal and cardiac troponin. Lindert et al. demonstrated that cardiac troponin, which does not open to the same fully open conformation as its skeletal counterpart on its own, without the presence of TnI[19], in fact undergoes an increase in the variability of structure, indicating a dynamics-based mechanism of troponin C opening rather than an absolute structural shift. [20, 21] Genchev et al. utilized a computational simulation to demonstrate similar, though not identical results using skeletal troponin, showing that without TnI, skeletal troponin quickly closed to a semi-open state rather than its full open conformation[22]. Building off of these computational studies is the primary focus of this investigation. While structural studies indicate a direct conformational change between closed and open states and exposure of the hydrophobic patch buried in troponin[23], studies like those performed by Genchev et al. suggest that this may not be the case, and instead suggest the possibility of a dynamics-based change like that which was found in cardiac troponin[21, 22].

# Chapter 2:

# Full Model Creation Using the

# Rosetta Suite

**Rosetta CM**

The number of fully solved protein structures is always growing. The Protein Databank currently contains over 120,000 different macromolecular structures[24]. However, there are still a vast number of structures that remain unsolved. For these gaps in scientific knowledge, it can be useful to utilize computational substitutes for experimental data. Rosetta is a full suite of software designed for the purpose of macromolecular modelling, from *de novo* structure determination to protein-protein docking. The Rosetta library includes a number of protocols written to most effectively utilize the power of the Rosetta software, and two are of particular significance in this project. The RosettaCM protocol is used for comparative modelling of protein structures[25], while the RosettaLigand protocol is used for docking small ligands to binding sites within a protein[26].

Comparative modelling is a method of structure determination that uses protein structures that have already been determined as the basis for creating a structure for a protein for which no model exists. While programs such as NAMD are almost entirely theoretical, utilizing only a series of algorithms to determine movement once the initial structures are provided, RosettaCM includes a knowledge-based approach. Previous experimental data is utilized extensively throughout the process in a number of ways. The general process for forming the model of a new protein is outlined below, as well as in Figure 2[25].

*Concepts and Implementation*

<u>Template threading</u>

The process begins with at least one template protein structure (in .pdb format) and the desired protein sequence. The target sequence and template sequences are aligned to match residues that are either identical or similar among multiple sequences. Once the alignment has been generated, the threading process can begin. A model of the target sequence is generated, and sections that are identical or similar are aligned in space so that backbone structure is conserved as much as possible, as shown in Figure 2[25].

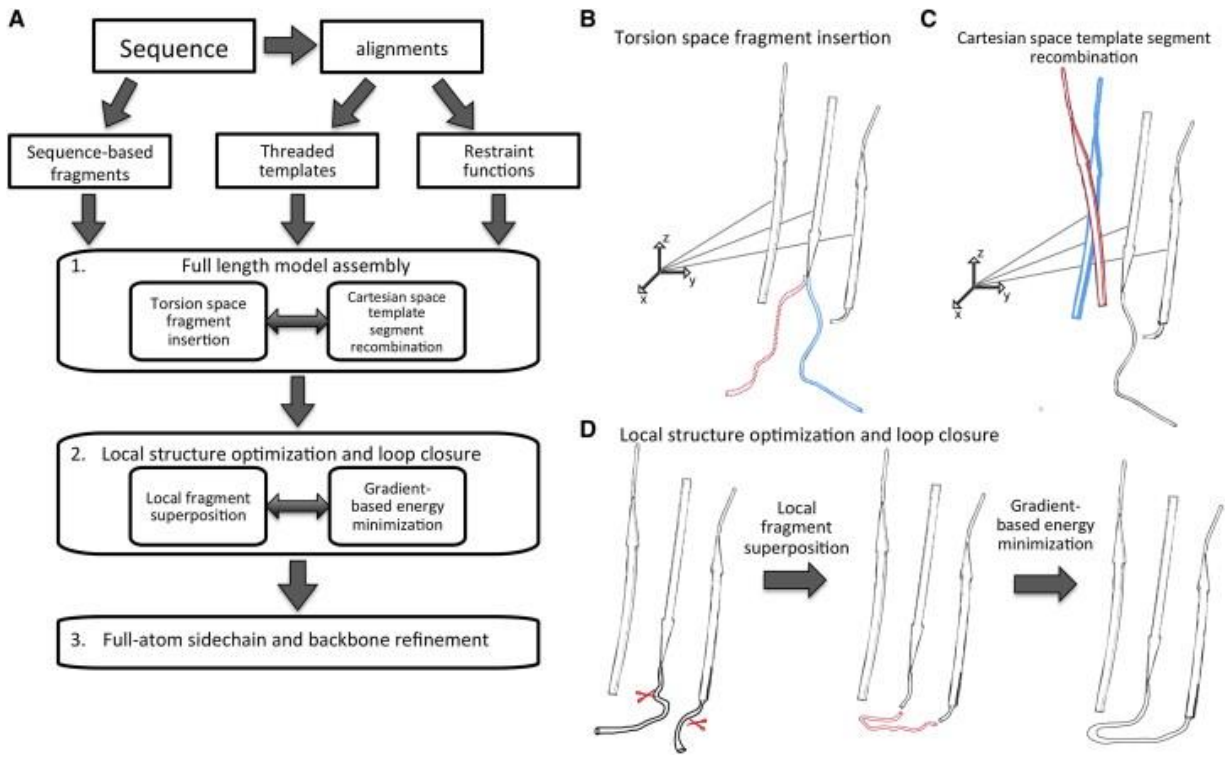



Figure 2: (L) A flowchart detailing the process of forming a protein model through comparative modelling (R) Illustrations demonstrating the methods described in the process[25]

*Full model assembly*

This threading process creates one model for each template provided, and while these models are very close to the template, they do not take into account the differences

between template and target, and they must also be combined into a single structure for use in molecular simulations. To accomplish this, Rosetta first splits each of the models into fragments (partial threads). This is done primarily by preserving secondary structure elements (helices and beta sheets) as whole segments, with loop regions between elements being split in half and combined with the attached secondary structures. This essentially leaves a pool of fragments from the various templates that can be recombined to build the full structure. To supplement this recombination, the sequence is also used to generate fragment files. A program known as Robetta is used to take the various amino acid combinations in segments of 3 and 9 residues and searches across known protein models for identical fragments[27]. From there, the torsion (phi psi) angles are determined for each known fragment.

Once all fragments have been assembled, Rosetta begins assembling various full models using a combination of template fragments and *de novo* fragments generated by Rosetta using the fragment torsion angles. This process is repeated several times utilizing a different series of fragments each time to assemble a set of models (1,000-10,000 different models)[25]. These models are then assigned a score using RosettaCM's built-in score function, which weighs a number of factors to determine what it believes to be the "best" model[28]. The score function utilizes the differences between the target structure and the templates (using RMSD values) across aligned regions, as well as the overall energy of the atoms present in the structure, to select the most optimal structure[28].

Structures at this state often have correct topology, but are often severely distorted internally, particularly at loop regions. Since each loop is split in half, with each

side being affixed to its corresponding secondary structure, the loop regions are often disconnected in space, as seen in figure 2D. To remedy this, Rosetta CM replaces these two loop fragments and superimposes a de novo fragment from the fragment pool. This fragment is then smoothed with a local energy gradient minimization step. The models at this stage are nearly complete. All that remains is to explicitly represent the side chains, which had been approximated up to this point using the Rosetta centroid functions, and optimize their orientation in space.

## Methods

There is no existing experimental structure for human skeletal troponin C. To assemble a reliable model for human skeletal troponin, a total of four separate templates were used. Each of these models was taken from the Protein Databank. The structures for apo skeletal troponin from *Gallus gallus* (domesticated chicken)[23] and *Melleagris gallopavo* (domesticated turkey)[29] were used as templates for the apo form. The calcium bound model was formed using structures of calcium-bound troponin from *Gallus gallus*[23], as well as that of *Oryctolagus cuniculus*(European rabbit)[30].

First, a basic alignment of amino acid sequence was performed to align each of these models with tnnc2[31], the translated amino acid sequence of human troponin C. Alignment was performed using the Basic Local Alignment Search Tool (BLAST)[32]. As can be seen in Figure 3, the alignments are very strong, with differences primarily existing in the terminal and loop regions, and % identity values greater than 95%. These alignment files, along with the PDB files of the six models, were fed into the Rosetta software and the 1tnnc2 sequence was threaded into the basic shape of these models, yielding two sets of two models of human troponin.

```
1tnnc2 --MTDQQAEARSYLSEEMIAEFKAAFDMFDADGGGDISVKELGTVMRMLGQTPTKEELDAIIEEVDEDGSGTIDFEEFLVMMVRQMKEDA
1trf   ----------AFLSEEMIAEFKAAFDMFDADGGGDISTKELGTVMRMLGQNPTKEELDAIIEEVDEDGSGTIDFEEFLVMMVRQMK---
1tnp   ASMTDQQAEARAFLSEEMIAEFKAAFDMFDADGGGDISTKELGTVMRMLGQNPTKEELDAIIEEVDEDGSGTIDFEEFLVMMVRQMKEDA
1tcf   ---TDQQAEARSYLSEEMIAEFKAAFDMFDADGGGDISVKELGTVMRMLGQTPTKEELDAIIEEVDEDGSGTIDFEEFLVMMVRQMKEDA
```

Figure 3: Alignment of the 1tnnc2 sequence to 1TNP, 1TRF, and 1TCF[23, 29, 30, 32]

The sequence was also fed into the website Robetta[28], which used its knowledge databank to generate a list of fragment angles. One list contained all known angles for

three amino acid-length fragments that matched the given sequence, and another that did the same for fragments that were nine amino acids long.

These models were then run through the "Hybridize" protocol[28] (see steps 1 and 2 from Figure 2), which combined parts of the two models with segments from the fragment files to form a series of full hybrid models. A total of 1000 models were generated, each with a unique combination of threaded model and *de novo* fragments. These models were scored using RosettaCM scoring function, and the best scoring model was selected for the next step.

To this point, the processes for the apo and $Ca^{2+}$-bound versions have been largely identical, with the exception of the base models used. However, an additional step was necessary to account for one significant difference: the presence of the calcium ions. While the Rosetta CM documentation claims that it can integrate simple ligands and ions [25], all attempts at doing so through the original protocol were unsuccessful. Instead, a different Rosetta protocol was used: RosettaLigand docking. RosettaLigand is a docking protocol for docking small molecules into protein binding sites[26]. RosettaLigand utilizes a stochastic procedure combined with simple energy minimization of side chains to determine a series of possible binding locations within a given protein structure[26]. 20 different combinations of calcium and side chain placement were generated for each binding site. By aligning these calcium-binding loops with experimentally determined structures in Chimera visualization software[33], the best binding site was determined, primarily through visual analysis of the placement of nearby acidic residues that were likely vital in coordinating the positively-charged calcium.

Once the calcium was docked into the newly generated human troponin models, both it and the apo form could be completed. The final “Relax” protocol [25] was performed on both forms, generating 100 models that were primarily concerned with full side chain optimization. These models were clustered based on their RMSD, and the lowest score from the largest cluster was selected, yielding the two final models.

## Results and Discussion

The final models can be seen in Figure 4. The interhelical angle between helices A and B is shown, which will be relevant in the analysis of later simulations. Additionally, a visualization of the hydrophobic patch in both forms is also shown. The increased exposure of the hydrophobic residues is clearly evident in this visualization.

Comparative modelling and *de novo* protein structure prediction are still relatively young fields, and many still widely prefer more traditional experimental methods for structural discovery. This particular case was an ideal one, with multiple proteins with largely identical function across species as well as highly similar sequence identity being used as templates. However, the value of such software should not be underestimated. The program was able to construct a new protein model utilizing a sequence for which no such model existed, demonstrating the increasing ability of computational biochemistry

While this software is incredibly powerful and capable of far more than was asked of it in this case, comparative modelling was not the focus of this particular project. The construction of these full human skeletal troponin C models was simply the first preparatory step in the analysis of the dynamics associated with its activity.

A.

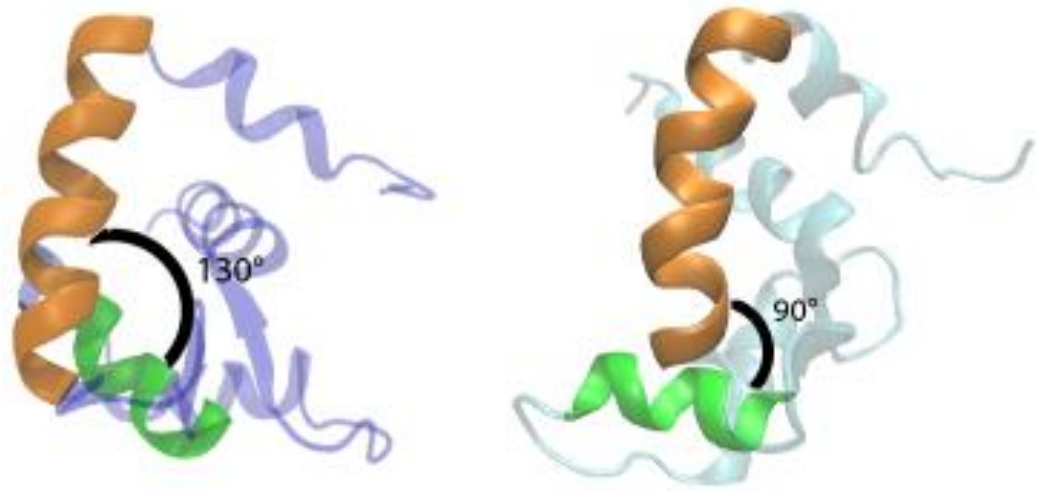
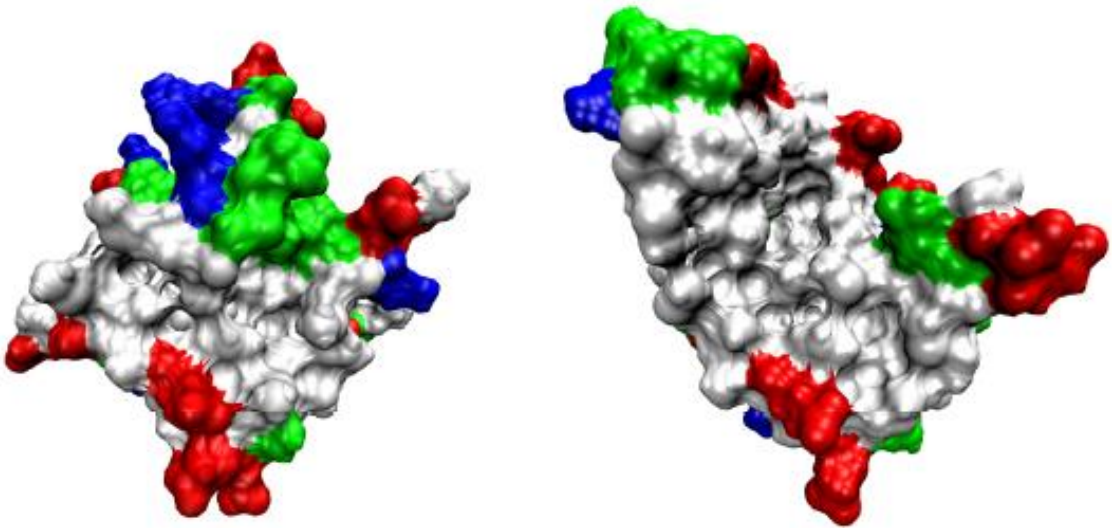



Fig 4a: Models of apo and calcium-bound Rosetta models, from L to R, with A-B interhelical angles indicated



4b: Van Der Waals visualization of the hydrophobic patch of the Rosetta model with apo on the left and calcium-bound on the right. Blue indicates positively charged residues, red indicates negatively charged residues, green indicates polar residues, and white indicates hydrophobic residues

# Chapter 3:

# Molecular Simulation of Troponin C with NAMD

**NAMD: Concepts and Application**

Due to the incredibly small scale of troponin C shifts, both in space and time, computational simulations are an ideal tool for studying this its mechanism. There are a number of different programs designed for such problems, each with their own method of effectively modeling the various interactions present with proteins contained within cellular conditions. NAnoscale Molecular Dynamics (NAMD) represents one of the most prominent and well-known of these programs[34].

NAMD allows for simulations of relatively large systems (>100,000 atoms) using the Newtonian equation of motion shown in Figure 5, which is used to calculate the motion of each individual atom within the simulation.

$$m_\alpha \ddot{\vec{r}}_\alpha = -\frac{\partial}{\partial \vec{r}_\alpha} U_{\text{total}}(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N), \alpha = 1,2 \ldots N$$

Figure 5: The Newtonian equation of motion utilized by NAMD to simulate the motion of each individual atom[34].

The potential energy for each atom, represented here as $U_{total}$, is itself comprised of five separate terms that correspond to the energies associated with inter atomic forces. Three terms determine the energy of bonded atoms: bond stretching, bond angle, and bond torsion (dihedral angles); while the other two cover the Van der Waal's forces and electric (Coulomb) forces between non-bonded atoms in space.[34]

Limitations and Solutions

Simulations for cellular conditions, while relatively large, cover only a small space compared to that of an actual biological cell. If simulations were allowed to continue with no restrictions on the boundary of the system, then the molecules would invariably spread out over time as the intermolecular forces drive them apart. It also means that less force will be applied to atoms on the boundary, which limits the applicability of these simulations to real-life events. To account for this, NAMD utilizes periodic boundary conditions. This concept treats the simulation as an infinitely-repeating series of identical systems that perfectly mirror the activity of the original, extending in all directions. When an atom crosses the boundary, an identical copy of this atom is then inserted into the opposite side of the system. Since the system repeats, the atom experiences the same forces when it crosses the boundary as it did before crossing. So long as the boundary leaves enough room for individual copies of the protein to come into contact with one another, the simulation can appropriately mimic the normal cellular conditions.

While the periodic boundary solves the issue of behavior at the edge of the simulation, it introduces another problem. The force experienced by each atom in the system is dependent on the forces exerted upon it by every other atom in the simulation to which it is not bonded. This concept, in combination with the infinitely repeating periodic system, makes calculation of Coulomb forces and Van der Waals forces computationally unviable over any useful time-scale. To effectively balance efficiency with accuracy for these simulations, two separate modifications are made to the calculation of Coulomb forces.

The first is a simple cut-off distance that can be set by the user for each simulation. Atoms more than a certain distance apart from each other are ignored in Van der Waals force calculations[34]. Since the force is inversely proportional to the square of the distance between the two atoms, at longer distances the force becomes so small as to be negligent in calculations. This helps mitigate the effects of the infinitely-repeating boundary, but force calculation remains non-trivial in systems with tens or hundreds of thousands of atoms. To optimize electrostatic computations, NAMD uses the periodic conditions along with the particle-mesh Ewald (PME) method to compute the full electrostatic potential for the system[35, 36].

The PME method utilizes a three-dimensional mesh overlaying the entire system. For each particle in the system, the charge is distributed over the grid points surrounding it, with weights assigned to each gridpoint relative to the distance from the actual point of charge (see Fig. 6)[34]. The full algorithms that are utilized for these calculations are relatively complex and beyond the scope of this project, but this method allows for the computation of the full electrostatic potential across the whole system using the periodic boundary conditions while still remaining feasible for nanosecond-scale simulations.

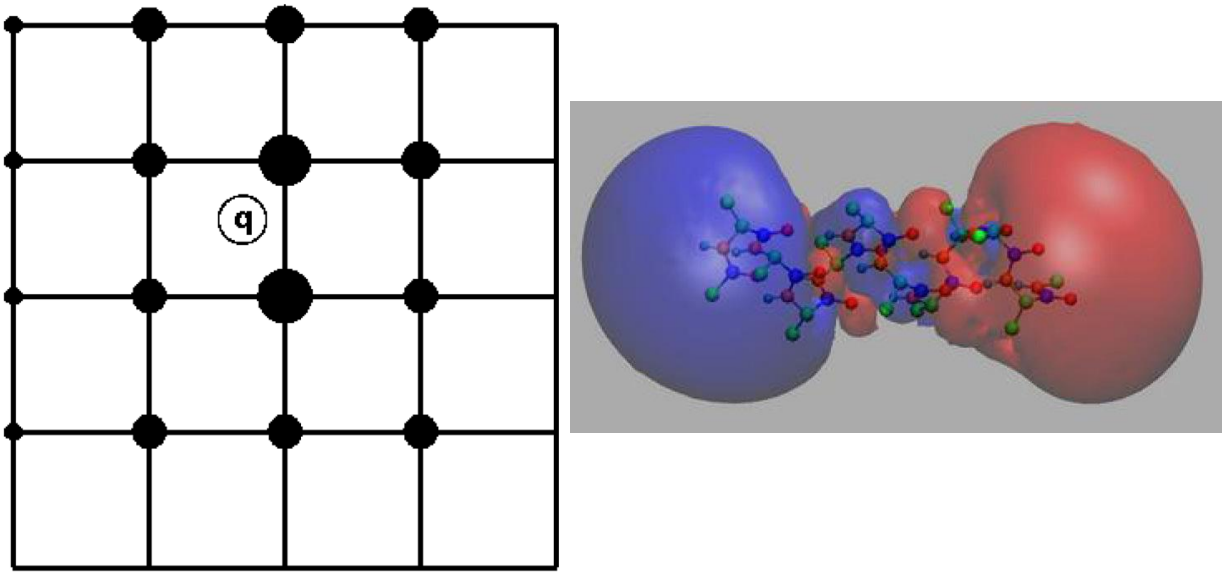



Figure 6. Left: 2-dimensional representation of the particle-mesh Ewald method, with charge *q* distributed across mesh points according to the distance to the charge. Right: A smoothed model of the electrostatic potential of decalanine calculated using the PME method (blue indicates positive charge, while red indicates negative)[34]

The power of this program is further increased by its multithread capabilities. NAMD possesses the ability to split its computational load into several parts, which can then be distributed over several processing units, or even a high-powered computational network[34]. This speeds up the process many times over and allows these large-scale simulations to be completed in a reasonable time frame.

## Methods

Once the models have been generated using the Rosetta suite, they must be modified somewhat to prepare them for simulations. From this point forward, the methods used to manipulate these two models are identical. The two models ($Ca^{2+}$-bound and apo forms) were first solvated by adding a box consisting of water molecules around the protein. This box was generated with edges 14 angstroms from the furthest atom on each side to allow for sufficient space for the periodic boundary conditions. The models are then ionized. A number of potassium and chloride ions are inserted into the water box. The number of ions gives a concentration approximately equal to cellular conditions. There are slightly fewer potassium ions contained in the calcium-bound model to account for the increased positive charge provided by the calcium. The resulting models contain ~20,000 atoms to be simulated between the protein itself, ions, and water molecules (Figure 6).

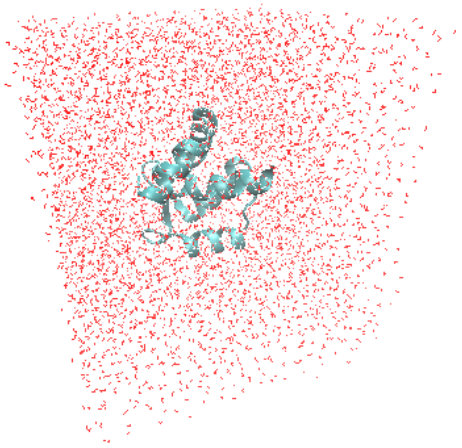


Figure 6: An image of apo troponin C, fully solvated within a large water box, tilted slightly at an angle.

While the systems now contain all molecules necessary for simulation, they are not quite ready for the full simulation runs. The water molecules, randomly placed around the protein, have not formed the hydration shell around the macromolecule as they would in an actual cell. To achieve this, the simulation begins with a minimization step. This step allows for energy minimization of water molecules situated around the protein.

To maintain the protein structure during this step, during the first part of the simulation, the atoms of the protein are held in place using a restraint file, which is used by NAMD to restrict the movement of the protein chain while the water and ions form around it. This minimization step (min_1) is run at 0 K to eliminate the movement of water molecules that occurs under normal conditions so that only the energy minimization is considered. Simulations were carried out using CHARMM force field parameters[37]. The first minimization step is run for 10,000 steps, with each step lasting 2 femtoseconds. Once this minimization step is completed, a second step is carried out for another 10,000 steps at 0 K, but without the restraint conditions. This allows for small minimization shifts in the protein in response to the formation of the water shell.

The system is now fully minimized, but still sits at 0 K. The system needs to run at 300 K to accurately simulate cellular conditions. To heat the system without significantly disturbing the contents, an equilibration step is necessary. This step pairs the cell with a simulated heat bath using Langevin dynamics[38], which introduces heat into the system while also accounting for frictional forces experienced by the collision of solvent molecules. This step again utilizes a restraint file, but slowly lessens the amount

of restraint it utilizes to allow for a gradual equilibration process that maintains protein structure. By the end of the equilibration, the models are ready for the full simulation.

The full simulation was completed in 100 ns increments. The Langevin heat bath was maintained at 300 K, with a second control implemented to maintain pressure. This control used a slightly modified version of the Hoover method[39], again using Langevin dynamics to control pressure fluctuations. The size of the cell is periodically adjusted to maintain an average pressure of 1.03125 bar (standard pressure at sea level). For this simulation, the timestep was set to 2 fs, meaning that the atoms moved for 2 femtoseconds under their current equation of motion before the simulation was momentarily paused to allow for a recalculation of the various forces. Every 1000 timesteps (2 picoseconds), the program saved a snapshot of the system, allowing for an accurate sampling of the various conformations that the protein experienced. The program also saves all information related to the velocity and position of the atoms in the system, allowing each subsequent run to begin where the previous one completed. 8 increments were run for each system, resulting in a total of 800 ns of simulation, with 40,000,000 frames saved for each.

**Results**:

The 800 ns of simulations for each model yielded 40,000,000 frames contained in 8 dcd files. Using the visualization software VMD[40], the partner program of NAMD, each of those frames was converted into its own separate PDB file, removing all waters and ions. This allows for analysis of the internal shifts of the protein conformation over time.

Interhelical Angle Analysis

It is difficult to quantify the "openness" of the hydrophobic patch to which TnI typically binds. To do so effectively, it is necessary to examine a certain interhelical angle. Experimental data shows that the angle between helices A (residues 14-27) and B (residues 37-46) varies greatly between "closed" and "open" conformations. Using a program known as interhlx [K. Yap, University of Toronto], these angles can be measured for each frame of the simulation

Apo form

The apo form of skeletal troponin C, as expected, remained largely unchanged throughout the simulation. As seen in Figure 7, the interhelical angle that signified the exposure of the hydrophobic patch rarely wavered, never dipping to the fully open 90º observed in experimental measurements of calcium-bound, open troponin C molecules (both skeletal and cardiac forms)[30]. In fact, the lowest angle experienced over the entire simulation was only 105º. This matches both expectations and previous experiments. Without the presence of calcium, it is expected that the troponin molecule would open incredibly rarely, if ever. If the protein were to open spontaneously, this would significantly affect the highly-controlled process of muscle contraction in humans.

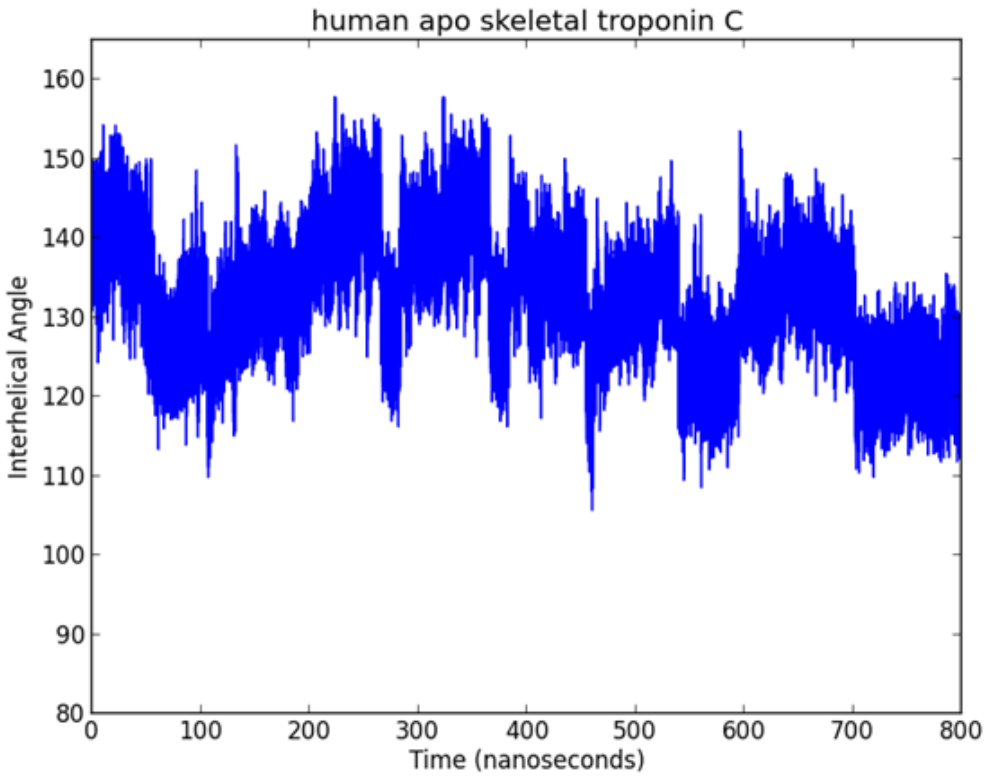



Fig. 7: Interhelical angle of apo skeletal troponin C helices B and C over time. Note the relative stability, with most frames falling between 110º and 150º. Compare to Fig. 8

<u>$Ca^{2+}$-bound form</u>

Unlike the apo form, the calcium bound form was highly variable throughout the time of the simulation, as can be seen in Figure 8. The interhelical angle began in the "open" range. However, within a few nanoseconds, the protein closed up, quickly reaching the typical angle of the closed conformation (~130º). Based on previous computational results[21, 22], this result was similar to cardiac troponin. This is somewhat

surprising given that calcium-bound models of skeletal troponin are found in a fully open state. While this similarity with cardiac troponin suggests that troponin C would remain open for incredibly short time fractions, the skeletal troponin seemed to settle in a much more open conformation, finding what seemed to be a separate equilibrium, perhaps due to an energy well of some sort, for a few nanoseconds. This more closely matches experimental results, but the shift was fairly temporary. Eventually, the protein once again reached a relatively closed conformation, with most angles falling between 100º and 140º. However, the protein did shift to the fully open conformation twice more within the simulation time, with hundreds of nanoseconds between each opening event.

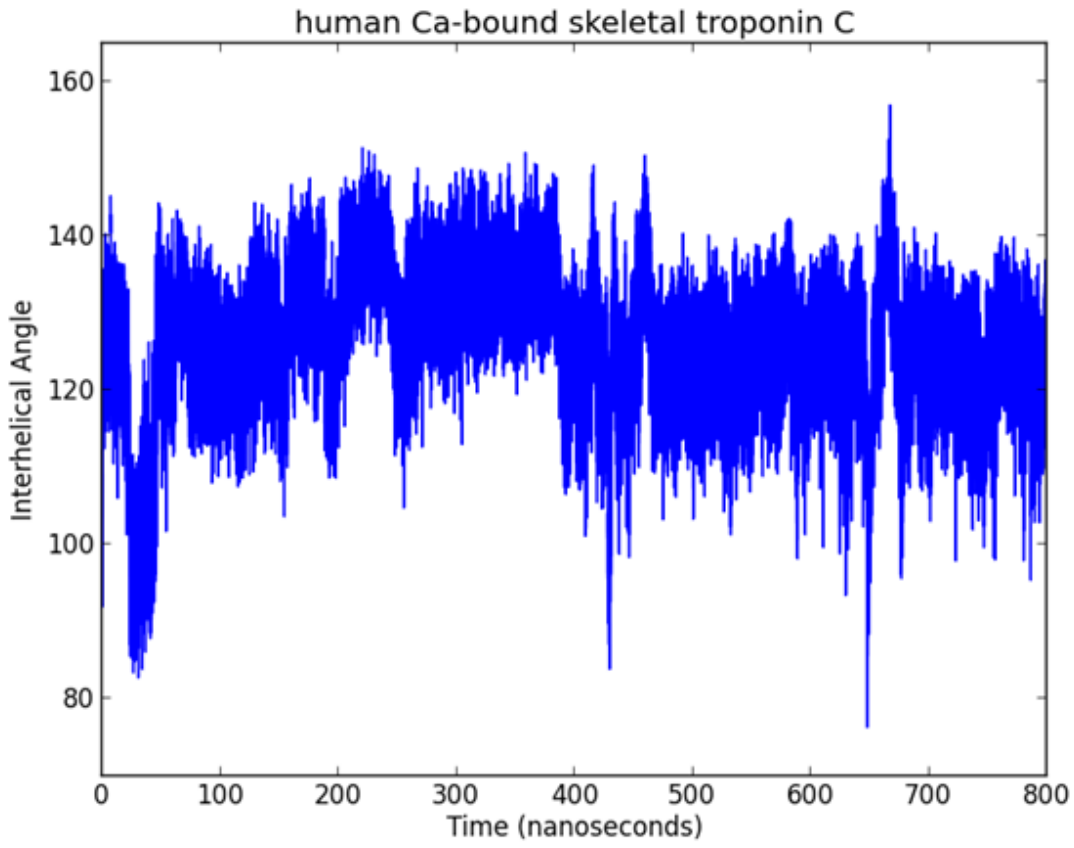



Figure 8: Plot of the interhelical angle of helices A and B over time for calcium-bound skeletal troponin C.

**Comparison and Analysis**

While it is useful to visually examine the graphical representation of this opening event, the data can also be used quantitatively. By assigning a cutoff angle to indicate the split between open and closed conformations and treating these conformations as two distinct states of the protein, one can assume that the two separate energy states form a Boltzmann distribution based on the number of frames found in each state[20]. Using this assumption and the equation for a Boltzmann distribution of states (Figure 9), the energy difference between open and closed states can be quantified. The results of these calculations are shown in Figure 10

$$\Delta G = kT \ln N_{\text{closed}} / N_{\text{open}}$$

Figure 9: Calculation of energy difference utilizing the Boltzmann distribution equation[20]

Based on this data, it is obvious why the open conformation was never observed for the apo form of TnC. A simple linear extrapolation of this curve would show the massive energy barrier that the system would need to overcome in order to shift the apo form to an open state.

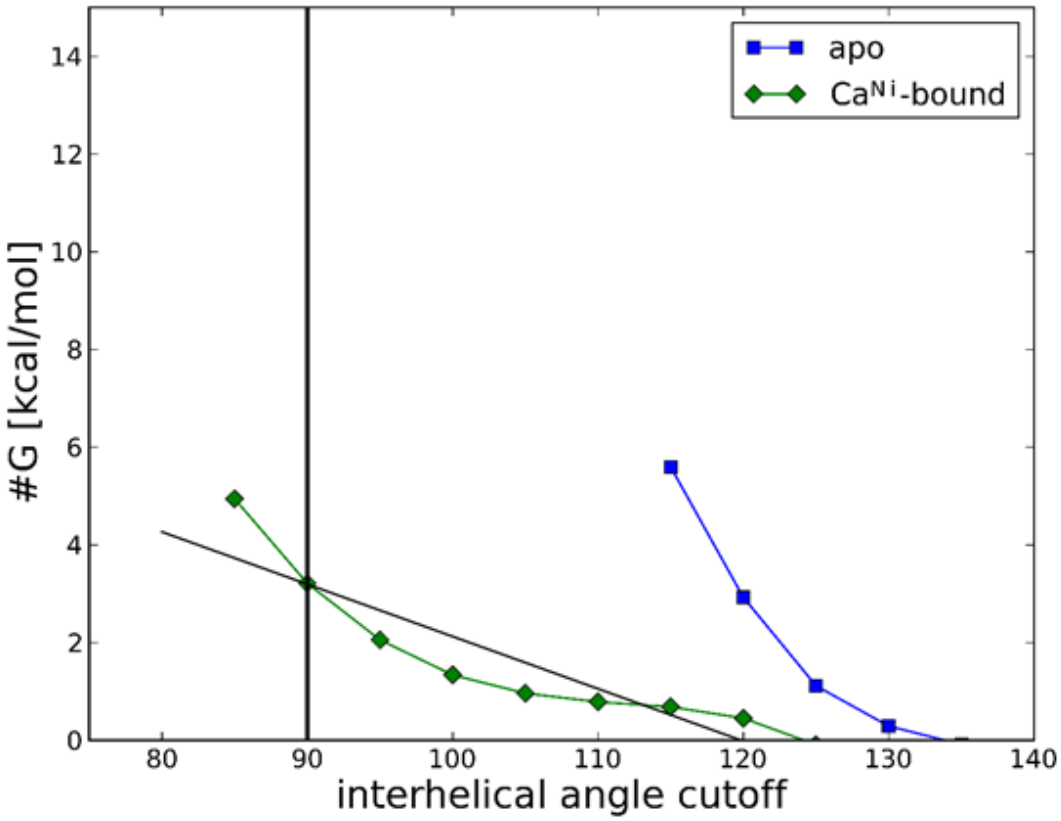



Figure 10: Energy barrier using various interhelical angles as the cutoff between "open" and "closed" conformations, calculated using the Boltzmann distribution of states equation.

For the calcium-bound form, this energy barrier is not nearly as insurmountable. Based on the traditional definition of open state as any protein with interhelical angle < 90º, the energy barrier between the open and closed conformations of calcium-bound troponin is only 3.2 kcal/mole. While this barrier is still significant, as evidenced by the fact that the calcium-bound form still spends most of the simulation in a relatively closed conformation, it is not nearly as massive as that of the apo form. While the calcium-bound form follows a relatively linear trend when comparing cutoff angle to energy

difference ($R^2$=.90), its linearity is somewhat diminished by the apparent dip in the trend between 95 and 110. This is likely due to the previously-mentioned state of near-stability for a few nanoseconds in the early parts of the simulation.

Similarly, this data can also be used to quantify the average time in between opening events by dividing the time spent in the open state by the number of times the transition between states occurred. The results of this calculation are shown in Figure 11. When graphed on a logarithmic timescale, the results show a similar trend to that of the energy change. The time between opening events is inversely related to the cutoff angle in a somewhat linear fashion. For all angles that the apo form exhibited, the apo form spent a significantly greater time between opening events, and based on other data, it can be expected that this trend would continue. Once again, the aberration in the mostly linear-seeming trend line is likely due to the early fluctuation near the open conformation that occurred in the calcium-bound state. Using the standard cutoff angle of 90º, skeletal troponin C spent ~3 nanoseconds on average in the closed conformation in between opening events. This may seem like a relatively short time-scale, but two things should be noted. First, even a 3-nanosecond gap is significant compared to the length of time that the molecule spends in the closed conformation, which is often on the order of a few picoseconds. Second, this gap includes the short amount of time that the protein spent fluctuating just above the cutoff angle. Note the distance between the other two significant opening events, which have nearly 200 nanoseconds between them.

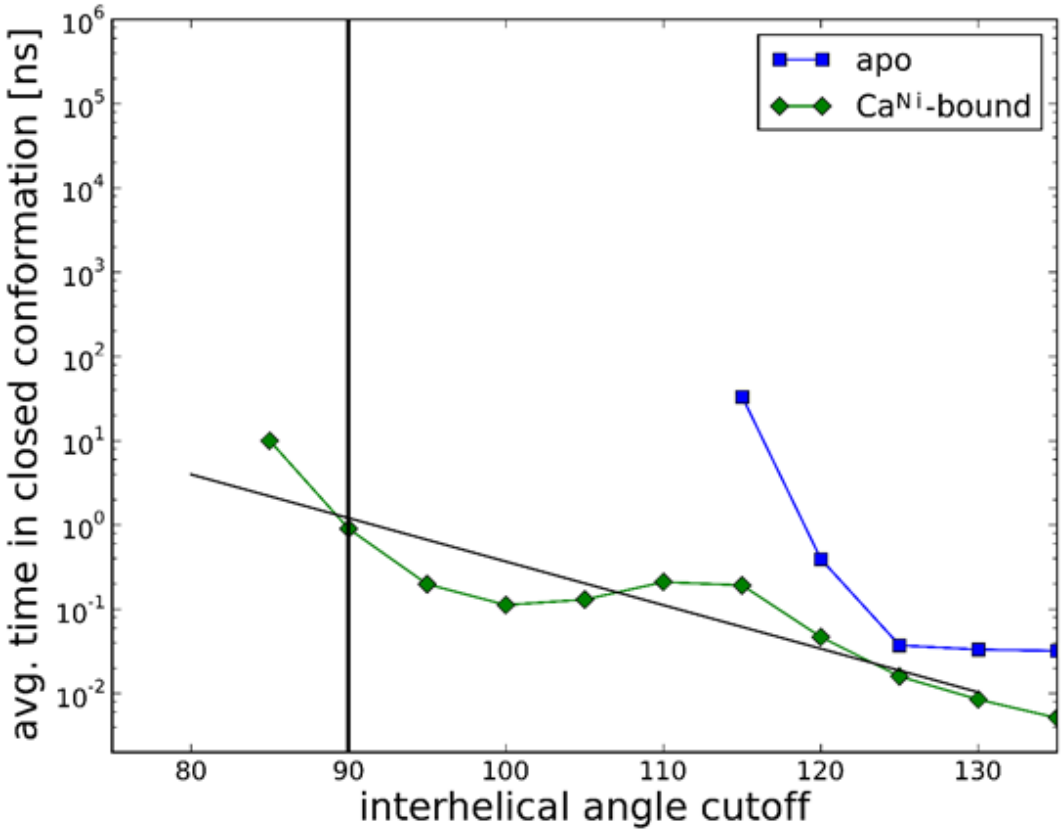



Figure 11: Average time in nanoseconds between opening events, based on different cutoff angles

Based purely on examination of the interhelical angle plot, one may assume that the calcium-bound form possesses a more dynamic nature. This assumption can be quantified in a number of ways using VMD, which can calculate the root mean square deviation (RMSD) of the various atoms within the protein. Using this information, a numerical value can be assigned to the level of fluctuation from the starting point across time, as shown in Figure 12. It is clear from this information that after an initial stabilization, the apo form remains relatively unchanged, with few major differences in its structure. The calcium-bound form, on the other hand, is demonstrably less stable, with a number of major spikes in RMSD, indicating a significant shift in protein structure.

There is clearly a significant increase in the dynamic nature of the protein upon the binding of calcium, based on this evidence.

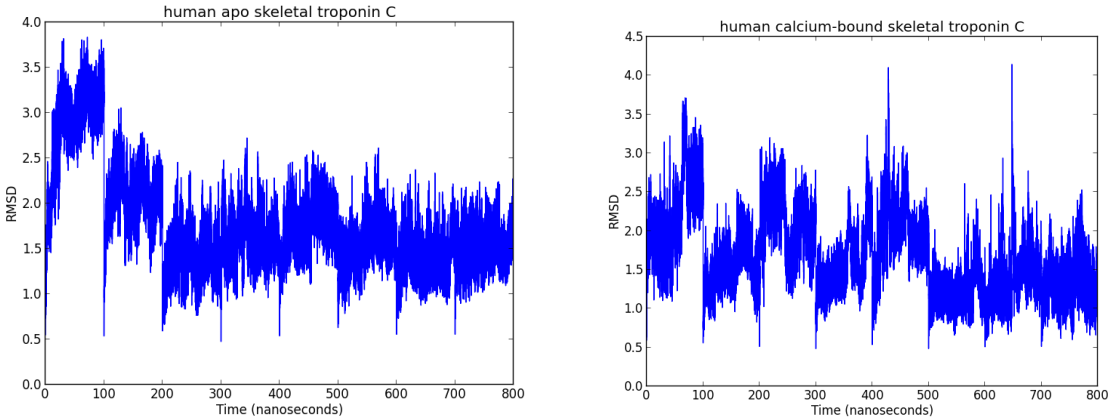



Fig. 12: Side-by-side comparisons of RMSD over time for both apo (L) and calcium-bound (R) states

While this evidence of increased fluctuation is useful, even protein subunits as small as troponin C are too large to have their behavior fully defined using a single value. VMD allows for another useful feature in its ability to calculate the RMSF, or root-mean-square fluctuation, which measures exactly how much each atom varies across the entire simulation. By measuring the alpha carbons of each amino acid, this measurement can provide an accurate picture of exactly which residues provide much of the variability in the structure. Comparing the two forms in Figure 13, it is clear that the residues share similar properties regardless of the presence of calcium. However, it is important to note the significant differences, particularly the difference in fluctuation between residues 20 and 50, which also contains the two helices that are of significant interest to this project. This provides further evidence in support of the idea that the binding of calcium induces a shift in the dynamics of the protein, causing it to become more variable in its structure. The spike in fluctuation between residues 60 and 70 is

also notable, but there is little indication that this portion of the protein has significant relevance in the hydrophobic patch opening event[22].

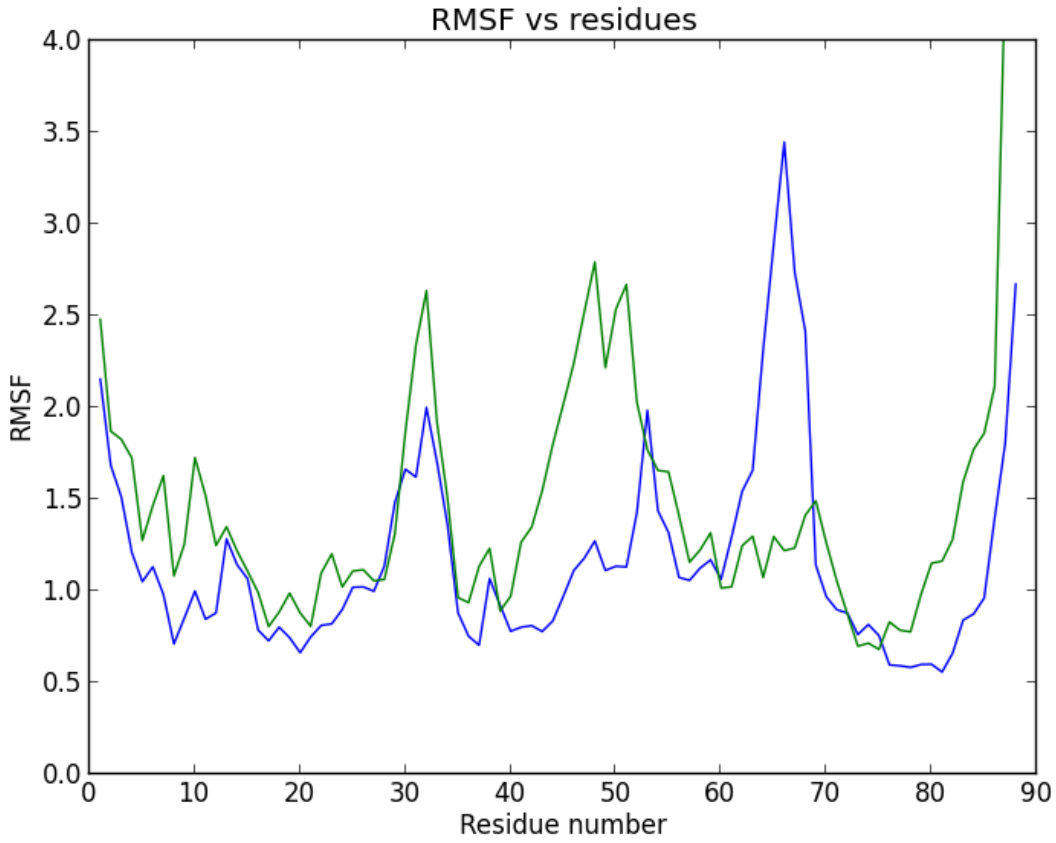



Figure 13: RMSF values for amino acids for both protein forms. Note the higher values present for the $Ca^{2+}$-bound form in the relevant regions indicating the second (B) helix and the loop between helices A and B (28-36, 37-46)

**Discussion**

Based on this data, it seems that calcium binding does not induce a direct conformational change that allows for exposure of the hydrophobic site of troponin C, but rather an increase in the variability of the structure. The energy barrier that the

protein must surmount in order to open is nearly insurmountable without calcium, and even with the calcium binding, it is still significant, on the order of 3-4 kcal per mole. Without TnI to stabilize the open structure, evidence suggests that even the calcium-bound version would spend much of its time in the closed conformation. This obviously does not agree with experimental NMR data, which shows skeletal troponin C in a fully open conformation when calcium is present. This computational data should not be taken as conclusive, especially in light of this conflict with experimental evidence, but it is certainly significant to note the significant energy difference between the two forms.

The increased dynamic nature of calcium-bound skeletal troponin C is also clearly evidenced in this simulation. Increased variability, particularly between the A and B helices of the molecule, is shown by the significant differences in the RMSD and RMSF values shown in these simulations. The calcium-bound form obviously samples a far greater number of conformations in the relevant EF had domain over time due to the decreased structural stability, and this sampling clearly trends towards the more open conformations, but the protein never fully settles into an open state. It instead remains highly variable throughout the life of the simulation. It is possible that the presence of TnI *in vivo* stabilizes the open structure through an induced fit mechanism, wherein the increased variability allows TnC to reach a state to which TnI can begin to bind, and this binding further stabilizes the most open conformation, leading to a more static open structure, but that is beyond the scope of this particular project.

# Chapter 4:

# Future Directions

These results lend themselves to a number of further questions. First and foremost is the disagreement between the results that show the closing of the hydrophobic pocket in skeletal troponin and the experimentally-determined structures. While the available NMR data suggests that the hydrophobic pocket remains open with relative consistency[30], this computational study shows a significant energy barrier that causes the protein to remain in a closed conformation for more than 99% of the time. It is worth investigating the reasons behind this discrepancy. Perhaps there is some interaction for which computational simulations fail to account. There are a number of approximations utilized to allow for simulation on a reasonable time-scale in NAMD[34], and it is possible that because of these methods, the integration of the equations of motion for these molecules does not include an interaction that stabilizes the open conformation of the molecule. It is even possible that peculiarities in the procedure used to clone and analyze the proteins structure led to conditions that altered the stability of the open conformation. This possibility has been proposed by Gallagher et. al. for structures determined via crystallography, where crystallization procedures led to possible stabilization of certain structures that did not completely agree with experimentation[41], but a possible mechanism for a similar discrepancy in NMR analysis is not known.

Regardless of the cause, it is clear that a difference does exist between this computational study and *in vitro* structural studies, and this discrepancy is an obvious avenue for further investigation.

Recent computational studies of troponin C have examined it in isolation. While this is generally more convenient, and can still give significant insight into the various

properties of the protein, it does not accurately represent the relevant interactions that take place within muscle fibers, particularly interactions with the TnI peptide. Studies by Lindert et al. and Genchev et al. have utilized some systems containing multiple components of the troponin system[21, 22], but computational studies of the protein-protein interactions remain relatively unexplored. Given the ever-increasing power of computation, research into these complex protein-protein interactions could take place in the near future.

Additionally, given the power of the Rosetta CM, it is theoretically possible to create models of mutations of the troponin protein. Given the number of studies that show the effects of various mutations on the activity and function of troponin[13-15], investigation of these mutations using computational methods would be a worthwhile avenue of study. If computational results continue to disagree with experimental trials as this investigation seemed to do, it could provide further insight into possible flaws in utilizing these numerical approximations of real-life physics in this particular context.

## References


1. Farah, C. S.; Reinach, F. C., The troponin complex and regulation of muscle contraction. *The Faseb Journal* **1995,** *9* (9), 755-767.

2. Squire, J. M., Architecture and function in the muscle sarcomere. *Curr Opin Struct Biol* **1997,** *7* (2), 247-57.

3. Lehninger, A. L.; Nelson, D. L.; Cox, M. M., *Lehninger principles of biochemistry*. 6th ed.; W.H. Freeman: New York, 2013.

4. Zot, A. S.; Potter, J. D., Structural aspects of troponin-tropomyosin regulation of skeletal muscle contraction. *Annu Rev Biophys Biophys Chem* **1987,** *16*, 535-59.

5. Lehman, W.; Craig, R.; Vibert, P., Ca(2+)-induced tropomyosin movement in Limulus thin filaments revealed by three-dimensional reconstruction. *Nature* **1994,** *368* (6466), 65-7.

6. Greaser, M. L.; Gergely, J., Reconstitution of troponin activity from three protein components. *J Biol Chem* **1971,** *246* (13), 4226-33.

7. Potter, J. D.; Gergely, J., Troponin, tropomyosin, and actin interactions in the Ca2+ regulation of muscle contraction. *Biochemistry* **1974,** *13* (13), 2697-703.

8. Herzberg, O.; James, M. N., Refined crystal structure of troponin C from turkey skeletal muscle at 2.0 A resolution. *J Mol Biol* **1988,** *203* (3), 761-79.

9. Leavis, P. C.; Rosenfeld, S. S.; Gergely, J.; Grabarek, Z.; Drabikowski, W., Proteolytic fragments of troponin C. Localization of high and low affinity Ca2+ binding sites and interactions with troponin I and troponin T. *J Biol Chem* **1978,** *253* (15), 5452-9.

10. Zot, H. G.; Potter, J. D., A structural role for the Ca2+-Mg2+ sites on troponin C in the regulation of muscle contraction. Preparation and properties of troponin C depleted myofibrils. *J Biol Chem* **1982,** *257* (13), 7678-83.

11. Johnson, J. D.; Charlton, S. C.; Potter, J. D., A fluorescence stopped flow analysis of Ca2+ exchange with troponin C. *J Biol Chem* **1979,** *254* (9), 3497-502.

12. Herzberg, O.; Moult, J.; James, M. N., A model for the Ca2+-induced conformational transition of troponin C. A trigger for muscle contraction. *J Biol Chem* **1986,** *261* (6), 2638-44.

13. Grabarek, Z.; Tan, R. Y.; Wang, J.; Tao, T.; Gergely, J., Inhibition of mutant troponin C activity by an intra-domain disulphide bond. *Nature* **1990,** *345* (6271), 132-5.

14. Fujimori, K.; Sorenson, M.; Herzberg, O.; Moult, J.; Reinach, F. C., Probing the calcium-induced conformational transition of troponin C with site-directed mutants. *Nature* **1990,** *345* (6271), 182-4.

15. da Silva, A. C.; de Araujo, A. H.; Herzberg, O.; Moult, J.; Sorenson, M.; Reinach, F. C., Troponin-C mutants with increased calcium affinity. *Eur J Biochem* **1993,** *213* (1), 599-604.

16. Ohtsuki, I.; Maruyama, K.; Ebashi, S., Regulatory and cytoskeletal proteins of vertebrate skeletal muscle. *Adv Protein Chem* **1986,** *38*, 1-67.

17. Grabarek, Z.; Tao, T.; Gergely, J., Molecular mechanism of troponin-C function. *J Muscle Res Cell Motil* **1992,** *13* (4), 383-93.

18. Farah, C. S.; Miyamoto, C. A.; Ramos, C. H.; da Silva, A. C.; Quaggio, R. B.; Fujimori, K.; Smillie, L. B.; Reinach, F. C., Structural and regulatory functions of the NH2- and COOH-terminal regions of skeletal muscle troponin I. *J Biol Chem* **1994,** *269* (7), 5230-40.

19. Li, M. X.; Spyracopoulos, L.; Sykes, B. D., Binding of cardiac troponin-I147-163 induces a structural opening in human cardiac troponin-C. *Biochemistry* **1999,** *38* (26), 8289-98.

20. Lindert, S.; Kekenes-Huskey, P. M.; McCammon, J. A., Long-timescale molecular dynamics simulations elucidate the dynamics and kinetics of exposure of the hydrophobic patch in troponin C. *Biophys J* **2012,** *103* (8), 1784-9.

21. Lindert, S.; Kekenes-Huskey, P. M.; Huber, G.; Pierce, L.; McCammon, J. A., Dynamics and calcium association to the N-terminal regulatory domain of human cardiac troponin C: a multiscale computational study. *J Phys Chem B* **2012,** *116* (29), 8449-59.

22. Genchev, G. Z.; Kobayashi, T.; Lu, H., Calcium induced regulation of skeletal troponin--computational insights from molecular dynamics simulations. *PLoS One* **2013,** *8* (3), e58313.

23. Gagné, S. M.; Tsuda, S.; Li, M. X.; Smillie, L. B.; Sykes, B. D., Structures of the troponin C regulatory domains in the apo and calcium-saturated states. *Nat Struct Biol* **1995,** *2* (9), 784-9.

24. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000,** *28* (1), 235-42.

25. Song, Y.; DiMaio, F.; Wang, R. Y.; Kim, D.; Miles, C.; Brunette, T.; Thompson, J.; Baker, D., High-resolution comparative modeling with RosettaCM. *Structure* **2013,** *21* (10), 1735-42.

26. Lemmon, G.; Meiler, J., Rosetta Ligand docking with flexible XML protocols. *Methods Mol Biol* **2012,** *819*, 143-55.

27. Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D., Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol* **1997,** *268* (1), 209-25.

28. Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D., Protein structure prediction using Rosetta. *Methods Enzymol* **2004,** *383*, 66-93.

29. Findlay, W. A.; Sönnichsen, F. D.; Sykes, B. D., Solution structure of the TR1C fragment of skeletal muscle troponin-C. *J Biol Chem* **1994,** *269* (9), 6773-8.

30. Slupsky, C. M.; Sykes, B. D., NMR solution structure of calcium-saturated skeletal muscle troponin C. *Biochemistry* **1995,** *34* (49), 15953-64.

31. Gahlmann, R.; Kedes, L., Cloning, structural analysis, and expression of the human fast twitch skeletal muscle troponin C gene. *J Biol Chem* **1990,** *265* (21), 12520-8.

32. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J Mol Biol* **1990,** *215* (3), 403-10.

33. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004,** *25* (13), 1605-12.

34. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005,** *26* (16), 1781-802.

35. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98* (12).

36. Ewald, P. P., Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physic* **1921,** *369* (3), 253-289.

37. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998,** *102* (18), 3586-616.

38. Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A Gen Phys* **1985,** *31* (3), 1695-1697.

39. Quigley, D.; Probert, M. I., Langevin dynamics in constant pressure extended systems. *J Chem Phys* **2004,** *120* (24), 11432-41.

40. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** *14* (1), 33-8, 27-8.

41. Vigil, D.; Gallagher, S. C.; Trewhella, J.; García, A. E., Functional dynamics of the hydrophobic cleft in the N-domain of calmodulin. *Biophys J* **2001,** *80* (5), 2082-92.